18th ICCRTS
C2 in Underdeveloped, Degraded and Denied Operational Environments

Title of Paper:

Controlled English for Effective Communication during Coalition Operations

Topics:

Data, Information and Knowledge
Collaboration, Shared Awareness, and Decision Making

Name of Authors:

Stephen Poteet; Ping Xue; Anne Kao
Research & Technology, The Boeing Company
P.O. Box 3707 MC 7L-43
Seattle, WA 98124-2207. USA
Email: ping.xue@boeing.com; stephen.r.poteet@boeing.com; anne.kao@boeing.com

David Mott; Dave Braines
Emerging Technology Services, IBM United Kingdom Ltd.
Hursley Park, Winchester, SO21 2JN, UK
Email: MOTTD@uk.ibm.com; dave_braines@uk.ibm.com

Cheryl Giammanco
US Army Research Laboratory
Human Research & Engineering Directorate
ATTN: RDRL-HRS-E
459 Mulberry Point Road
Aberdeen Proving Ground, MD 21005-5425
Email: cheryl.giammanco@usarmy.mil

Point of Contact:

Stephen Poteet
stephen.r.poteet@boeing.com
(425) 373-2783

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUN 2013		**00-00-2013 to 00-00-2013**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Controlled English for Effective Communication during Coalition Operations	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **The Boeing Company,Research & Technology,P.O. Box 3707 MC 7L-43,Seattle,WA,98124-2207**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

Presented at the 18th International Command & Control Research & Technology Symposium (ICCRTS) held 19-21 June, 2013 in Alexandria, VA.

14. ABSTRACT

Coalition operations involve multi-team and/or multi-nation collaborations. Linguistic variations and cultural differences often create unexpected challenges for effective communication and thus for Command and Control (C2) during military operations. In this paper, we propose using a controlled natural language, namely International Technology Alliance Controlled English (CE), and CE-based tools to improve cross-linguistic/cross-cultural communication. We will discuss various types of linguistic variations and cultural differences manifested by US and UK groups during coalition operations. The differences include not only lexical differences but more importantly differences in language use. These differences often result in miscommunication and impede effective operations. CE (Mott 2010) is a subset of English with a restricted grammar based on a formal syntax and semantics, which is human friendly but allows machine processing. The current version of CE provides a common form of expression that ? promotes standard terminology and usage to reduce ambiguity in person to person communication ? allows end-users to create new concepts with associated syntax and semantics ? provides a basis for automated and assistive applications and tools that support natural human-computer interaction, reasoning, and explanation CE and CE-based tools can play an important role in facilitating cross-linguistic and cross-culture communication and enabling multi-nation teams to work together effectively and efficiently.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**36**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Abstract

Coalition operations involve multi-team and/or multi-nation collaborations. Linguistic variations and cultural differences often create unexpected challenges for effective communication and thus for Command and Control (C2) during military operations. In this paper, we propose using a controlled natural language, namely International Technology Alliance Controlled English (CE), and CE-based tools to improve cross-linguistic/cross-cultural communication.

We will discuss various types of linguistic variations and cultural differences manifested by US and UK groups during coalition operations. The differences include not only lexical differences but more importantly differences in language use. These differences often result in miscommunication and impede effective operations.

CE (Mott 2010) is a subset of English with a restricted grammar based on a formal syntax and semantics, which is human friendly but allows machine processing. The current version of CE provides a common form of expression that:

- promotes standard terminology and usage to reduce ambiguity in person to person communication
- allows end-users to create new concepts with associated syntax and semantics
- provides a basis for automated and assistive applications and tools that support natural human-computer interaction, reasoning, and explanation

CE and CE-based tools can play an important role in facilitating cross-linguistic and cross-culture communication and enabling multi-nation teams to work together effectively and efficiently.

Keywords: Coalition operations, multi-nation collaborations, linguistic variations, cultural differences, cross-linguistic and cross-culture miscommunications, Controlled English, knowledge sharing, effective communication, situation awareness, decision-making.

1. Background and Needs

The US and the UK have established a collaborative research alliance called *International Technology Alliance* (ITA)[1] to address problems and challenges the coalition forces face during military operations. One of the major challenges is related to the fact that current coalition operations involve multi-nation collaborations with groups

[1] In 2006, the US Army Research Laboratory (ARL) and the UK Ministry of Defence (MoD) established the ITA as a collaborative research alliance with academia and industry partners to address fundamental issues in *Network and Information Sciences* to enhance the abilities of the US and UK to conduct coalition operations. The ITA is a unique UK-US collaborative venture. It is a multi-disciplinary research program that focuses on coalition needs and seeks to develop a mutual understanding and strong US-UK partnerships among the government, academia and industry participants.

and team members from diverse backgrounds (Pierce 2002a, 2002b; Chiarelli and Michaelis 2005). Efficient and successful collaboration requires effective communication and information sharing among coalition partners so that the related parties have a common understanding of goals and objectives. This is not easily achievable in the context of coalition operations. During multinational coalition operations, linguistic variations and cultural differences among multinational groups and team members have presented serious challenges in coalition communication and have been recognized within ITA as one of the Grand Challenges (Verma 2009).

In this paper, we discuss an approach to improving cross-linguistic and cross-culture communication. More specifically, we discuss the role that a controlled natural language (CNL) and CNL-based tools can play in reducing or minimizing miscommunication and enabling multi-nation teams to work together more effectively and efficiently. We will focus in particular on issues of language variation and differences in conceptualization across domains. Section 2 reviews the issues of miscommunication among multi-national groups and team members. In particular, it focuses on potential communication issues between US and UK personnel working in a coalition environment. Section 3 discusses a CNL implementation based on International Technology Alliance Controlled English (ITA CE, hereafter CE). Section 4 discusses how CE can be used to support human-to-machine interactions and human-to-human communications. Finally, section 5 summarizes CE and its utility for facilitating cross-linguistic and cross-cultural communications. It also discusses future work in extending CE capabilities and coalition applications in support of effective communications, knowledge-sharing and decision-making.

2. Miscommunication in Coalition Operations

Poteet et al. (2008a, 2008b, 2009) studied the miscommunications in coalition operations, particularly the linguistic aspects of miscommunications between English speaking US and UK military forces. It analyzed the relation between cultural differences and variations of language use, and its impact on miscommunication. The results of the analysis supported two initial hypotheses:

- Various types of linguistic differences exist at various levels of language use between British and American which lead to misunderstanding.

- Cultural differences result in variations in language use even though the US and UK share a ‘common’ English language.

These differences often result in outright miscommunication or otherwise impede effective communication. More specifically, the studies identified several aspects of language and language use that reveal linguistic variations between the US and UK groups that led to communication problems. These include:

- Use of Acronyms:

  Acronyms are commonly used among the military communities. Acronyms can pose a problem because they usually originate from a specific technical or culture group and are not known by people outside that group. For example, British military acronyms ("SO1" for Staff Officer 1, "SO2" etc.) are unknown to most American military personnel.

- Use of Slang, Colloquialisms and Jargon

  Use of slang, colloquialisms and jargon are related to the notion of language register or style. People use appropriate styles depending on the settings in question. The problem is that slang, colloquial expressions and jargon typically belong to a specific community or group, whether national, social, economic, organizational, or technical. While they allow for efficient intra-group communications, they are more likely to be misunderstood when used in a coalition setting, where people from other groups are involved. For example, people working night operations use specialized equipment (e.g. night vision equipment) and may use their jargon to describe things in terms that are unintelligible to the daytime operators who are not familiar with the equipment.

- Denotation vs. Connotation

  The intended or implied meaning of a sentence is often ambiguous, relying on the context in which the sentence is used. There is often a difference between what an expression literally denotes and what it is intended to suggest. The interpretation of the later relies on the understanding of the context. In this regard, miscommunication is more likely to occur in a coalition environment for where the different groups involved have different linguistic and cultural backgrounds. For example, in response to a misunderstanding of protocol, a UK warfighter stationed in the US and serving in a US unit said that they should come up with an "Idiot's Guide" for that situation. In the UK, "Idiot's Guide" refers to a primer type introduction to some topic, what in the US would be referred to as a "Dummy's Guide". The person he had emailed this to felt his intelligence was being impugned because of the negative connotation of "Idiot" and it took some time to sort out the misunderstanding.

- Misinterpreted Speech Act

  Speech acts refer to the various types of function that language can perform. In addition to making statements, language can be used to ask questions, make promises or requests, order, greet, etc. Correct interpretation of a sentence will depend on correct interpretation of the speech act of the sentence. An interrogative sentence such as "Can you pass the salt?" sounds like a question about the listener's capabilities but is more likely to be a request depending on the context. American English and British English show differences in terms of speech acts. For example, in British English, officers tend to give commands in a

form that sounds like a suggestion to American ears ("you might wish to …") and can be misinterpreted as such.

In short, language variation in a multi-national coalition setting can pose serious communication problems. Even for English-speaking teams, the English language used by team members from different nations (such as the US vs. the UK) and/or from different organizations may vary to some degree in vocabulary, sentence structure, language usage and style.

For improving communications between coalition forces, Poteet et al. (2008a, 2008b, 2009) made a number of recommendations including cross-linguistic and cross-cultural training to enhance cross-linguistic and cultural awareness. It would be most beneficial to have coalition partners train together prior to operations. However, this might be unrealistic due to time and other constraints.

It is also recommended that a standard terminology and a standard (or neutral) style or register be used in the coalition environment. While standardization of terminology and language use seems to be a potentially useful and important strategy to reduce ambiguity and thus avoid miscommunication, there are various practical issues involved in achieving this, as discussed in Poteet et al. (2008a, 2008b, 2009). One problem is enforcing usage of the standardized language. Another is the fact that the language will need to be constantly revised, since the situations encountered in the field are constantly changing. Experience suggests that one necessary component for supporting use of standard terminology and language is automated tools that enable access to standardized terminology and reinforce consistent language use, but also allow extensibility in an ever changing situation.

The central idea of our proposal is that CE can serve two purposes:

1. as a common, extensible standard language, supported and reinforced by CE-based tools, and
2. as the basis for communication tools (or add-ons to communication tools, like email and chat) that can recognize potential sources of miscommunication like those described above and alert users to their presence

CE can serve an important role in bridging linguistic, cultural and conceptual differences, and be useful in facilitating cross-domain knowledge sharing for effective communication between coalition partners.

3. CE and CE-based Coalition Sharing Capabilities

CE is a subset of English using a restricted set of grammar rules and a restricted vocabulary. Based on a formal syntax and semantics, CE provides both easy readability for human consumption and unambiguous representation of information for machine processing. Specifically it provides:

(i) A user-friendly language in a form of English, instead of, for example, a standard formal query language (e.g., SPARQL or SQL), which enables the user to construct queries to information systems in an intuitive way
(ii) A precise and formal language that enables clear, unambiguous representation of concepts and their relationships so that it can be used to build, extend and refine domain models by:
    a. Adding or modifying entity, relation, property, and situation types
    b. Describing how these concepts can be expressed in natural language (e.g. US English, or "american", or UK English, or "english")
(iii) An intuitive means of configuring the processing and reasoning that a CE-based system could perform

CE encourages a richer integration between human and machine reasoning capabilities in that it is human understandable yet machine readable. CE is designed to provide a human-friendly representation format for use by domain specialists (e.g. military planners and analysts) that may lack expertise in programming languages (Mott 2010, 2009; Mott et al. 2010; Mott and Hendler 2009).

3.1. CE Syntax and Statements

Some simple examples are given below to briefly introduce the CE syntax[2]. The creation of the domain model (or a general model across domains) using CE is accomplished by the definition of (domain) concepts, relationships and properties. These are all achieved through the "conceptualise"[3] statement. After a conceptualise statement had been made, the concept in question has been created within the CE domain model and statements relating to that concept can be made:

```
conceptualise a ~ person ~ P.
there is a person named Fred.
```

A slightly more advanced example would be:

```
conceptualise a ~ person ~ P that is an agent.

conceptualise the person P
        ~ is married to ~ the person P2 and
        has the value A as ~ age ~.
```

Thereby creating "person" as a sub-concept of "agent" and indicating that it can have the property of "age" and enter into a "married" relationship with someone, allowing the statement:

[2] See Mott 2010, 2009; Mott et al. 2010; Mott and Hendler 2009 for details of CE syntax and semantics.
[3] The UK spelling of "conceptualise" is due to the origin of CE at IBM, UK, although the language is extensible and does support the Americanized "conceptualize" also.

the person Fred is married to the person Jane.

and

the person Jane has 54 as age.

Note that the conceptualise statement does not force one to include all the relations and features it introduces in a single sentence; one's age and married status can be given in separate sentences. Rules may also be specified to provide further semantics of the concepts. For example the following rule expresses an important aspect of marriage:

if (the person P is married to the person Q)
then (the person Q is married to the person P).

The examples given so far, while simplistic, show how CE can be extended by adding new concepts and associated relations and properties that are relevant to the domain in question and can be as detailed or abstract as is appropriate to achieve the task at hand.

3.2 Mapping between Domain Concepts

CE provides a common mechanism to represent the relevant concepts for a domain. CE does not impose any constraints on the use of words for a specific concept or relation. As we discussed above, linguistic variants for the same or similar concept are common across domains and cultural communities. For example, the UK "petrol" versus the US "gas". On the other hand, the identical word or phrase may be used to represent somewhat different concepts. For example, the term "brigade" used among US military communities has subtly different meaning than that represented by the same term used in UK military communities – while it represents the same level in the hierarchy, it differs markedly in terms of number of personnel and amount of equipment and other resources.

We envision an aid to communication mediated by electronic means that will be aware of how natural language (e.g. words and phrases) maps into concepts and how different concepts are related to each other. Not only can different dialects or the jargon associated with different domains use different words to describe the same concept, but they may have slightly or radically different conceptualizations of the same situation. CE allows words and phrases (as well as more complex syntactic structures) to be mapped into concepts and CE sentences using these concepts, as well as rules relating different concepts.

In order to map between different domain models, we envision mapping between concepts instead of directly mapping between words. Knowledge sharing across cultures and domains is challenging. Different but related domains overlap but also differ to some extent in concepts and terminology. A common model is necessary, which is an aggregation of all the concepts and terminology of the related domain models as well as the mapping relations between those that are related but different (Giammanco et al. 2013). In reality the "common model" is likely to be a federation of separate but related models that span the domain in question and interlink/overlap, with mappings defined in such cases.

One needs to define the general concepts in question as well as record the specific instances of these concepts. For example, in the simplest case where different words are used for the same concept in UK English and US English, the human user would first need to do something like the following:

conceptualise a ~ common model concept ~ U that
    has the value V1 as ~ english word ~ and
    has the value V2 as ~ american word ~.

Then, to handle the petrol/gas example, they can say:

there is a common model concept named 'petroleum based fuel' that
    has 'gas' as american word and
    has 'petrol' as english word.

Everyone else in the system can offer their own mapping instances like “chips vs “crisps” etc. Another user might extend the model to add a new property of “canadian word” and another user may extend the model to add a feature like “severity” to reflect that some of these might cause serious problems and others would have less of an impact.

In somewhat more complex cases where two words in the two languages (whether the same or different) map to slightly different concepts, one would need to define the common model concepts they each need to be defined in terms of. For example, to handle the “brigade” example, one would need to define something like:

conceptualise a ~ military unit ~ MU that
    is an organization.

conceptualise a ~ UK brigade ~ that
    is a military unit and
    has ‘UK division’ as the next higher echelon and
    has N1 as ~ number of personnel ~.

the english word ‘UK brigade’
    expresses the concept UK brigade and
    has “brigade” as text.

conceptualise a ~ US brigade ~ that
    is a military unit and
    has US division as the next higher echelon and
    has N2 as ~ number of personnel ~.

the american word ‘US brigade’
    expresses the concept US brigade and
    has “brigade” as text.

Additional information would be encoded in the conceptualise statements for the two types of brigades in a real system and all the supporting concepts like "echelon" and "personnel" would need to be defined. We would also probably need rules to describe the difference and relationships between these two concepts. However, this gives an idea of how it would work. This captures the fact that "UK brigade" and "US brigade" are both military units and both have "brigade" as the text expressing them in their respective dialects.

Importantly, this CE-based approach is easy enough for the user to use and precise enough for the machine to process and interpret without ambiguity. Furthermore, because it is CE and based on an extensible model in a human language, it is plausible that such a system could be incrementally improved with lots of small updates by a wide community of users as experience is gained and lessons learned, rather than being dependent on a slow, fixed knowledge model upgrade cycle.

3.3. CE Inferencing and Rationale

Inference is the act or process of deriving logical conclusions based on known premises. CE builds upon earlier work on Controlled Natural Languages, such as Common Logic Controlled English (Sowa 2007) and aims to provide a single standard language for representation of all aspects of the information representation and reasoning space. CE is given a semantics by mapping onto First Order Predicate Logic. In addition to more traditional areas such as knowledge or domain model representation and corresponding information, CE also encompasses the representation of logical inference rules, rationale (reasoning steps), assumptions, statements of truth (and certainty) and has been used in other areas such as provenance and argumentation.

Rationale within CE is the formal explanation of the inference steps that were taken to reach a conclusion (Mott et al. 2010). The rationale information is also expressed in CE and may contain information about assumptions and true/false support pathways. CE inferencing and rationale capabilities can help resolve semantic ambiguities in sentences by representing and interpreting lexical and domain information (Mott and Poteet 2012). We hypothesise that it may also help to resolve pragmatic ambiguities such as those regarding intended meaning vs. literal meaning as discussed above. Displaying the rationale for the interpretation of a questionable or confusing statement or conclusion would make it clearer to the user what senses of the words and what interpretation rules were involved in reaching that conclusion, and having the rationale information available in the same CE format for machine processing can enable automated assistance in assessing rationale, for example in looking for patterns or trends.

4. CE System and Other Potential CE-based Systems

The aim of CE is to provide a common form of information representation that can be used by all parties, with different (but consistent and overlapping) domain models supporting each specialization and the whole endeavor. We have developed a system

with a set of underlying capabilities, which we refer to as "CE Store" that can be used to support some of the requirements of domain specialist users.

4.1 CE Store

The CE Store provides a basic CE processing and development environment[4] that includes the following high-level capabilities:

(i) Basic CE sentence parsing
(ii) Define/extend any concept model
(iii) Assert any CE sentence conforming to the appropriate conceptual model(s)
(iv) Define and execute CE queries including an example "visual query composition" element
(v) Define and execute CE-based logical inference rules, in the form of a "query with conclusion clauses" that can be used to assert new CE information
(vi) Define and execute "CE agents" in the form of Java code which conforms to a simple "CE Store" interface
(vii) Operate entirely in memory, or persist information to files
(viii) Some capability to convert to/from OWL and RDF formats

The CE Store also includes:

(i) An example of a web-based client to allow rapid development and browsing of CE-based information
(ii) Sample agents (written in Java and configured with CE) to carry out basic information processing tasks in the CE environment

CE and CE-based applications are designed to be most useful in situations that have the following characteristics:

(i) A high degree of human interaction, usually involving domain specialist users with complex needs in non-trivial environments.
(ii) A likelihood of rapidly evolving or uncertain tasks, queries or other knowledge-based activities.
(iii) The need for collaboration, either between different groups or team members, and/or across different disciplines.

Applications can be built in this environment using CE and can also include other non-CE based elements (e.g. maps and other visual interfaces) within the human computer interface. CE can be used to directly input general and specific facts and rules, but also is used by the system to display back to the user information that has been input by other means, e.g. via a graphical interface like that for building queries and rules.

[4] An alpha version of the CE Store (known as the IBM Controlled Natural Language Processing Environment) is publically available for download, at http://ibm.co/RDIa53

Our approach to lexical-conceptual capture and knowledge building relies upon the linking of words to concepts, or specific domain words to common model concepts. However, whereas the meaning of natural language words is generally understood by the community of speakers, the authoritative meaning of the concepts is only known to the specialized domain user who developed the conceptual model. For example, only the analyst can determine the linking of words to the concepts, although they can be assisted by tooling to perform this task. To this end we are developing an "Analyst's Helper" to assist the analyst in constructing the linguistic mappings between words and each concept in the conceptual model. To reduce the burden on the analyst, the Analyst's Helper leverages on-line resources like WordNet (Miller 1995; Fellbaum, 1998) to suggest possible words for each concept.

We have developed a model of language processing based on linguistic principles (Mott 2011) and have used this to perform NL processing and fact extraction, configured by knowledge represented in CE (Mott et al. 2012) and using the CE store. This extended representation of lexical information (grammatical and semantic) will aid more complex handling of the dialect relationships noted above, for example the matching of verbs and adjectives, and more complex descriptions of entities, to associated terms and concepts.

Once the knowledge discussed above about the relationship between UK English and US English has been entered into the system via the CE Store and the Analyst's Helper, software agents built in CE Store can take advantage of this knowledge. For example, by scanning chat in real time or looking in emails and documents and for communications between UK and US people it can alert the user to the use of ambiguous terms or terms with different meanings in various dialects or domains. The actual alert could take advantage of highlighting or color-coding to indicate such words in the text, but the precise meaning of the alternatives could be made evident by displaying the underlying CE sentences defining them or stating their relationship.

While this is an extremely trivial example, this approach could be used to handle the cases of acronyms, slang, and jargon discussed above. It could also help with terms that have alternatives in the different dialects, like "Idiot's Guide" and "Dummies Guide" by alerting to the existence of another version of the associated common model concept in another dialect. In addition, the basic pattern scales up well to more realistically complex cases, such as words with overlapping meaning, by indicating what they have in common and what features distinguish them.

More pragmatic or usage based problems like misinterpreting commands as suggestions would be more difficult to handle, but it is not only words and phrases that can be mapped, but potentially entire structures in the form of linguistic frames, although we have not looked into this very carefully yet.

The model and the data can easily be provided by humans and applied by software agents without the need for technical formats. In addition, the software agents to do the analysis and the alerting can be configured by CE in our existing system.

5. Conclusion and Future Work

Multi-team and multi-nation collaborations in coalition operations involve conceptual as well as terminological and other linguistic variations across domain models, which pose major challenges for cross-domain communication and information sharing among teams for efficient C2 operations. In this paper, we have discussed CE and CE-based tools that provide a common language and a platform for knowledge building in a standard form, which allows cross-domain knowledge sharing and aspires to encourage human-machine interaction. This will help members of coalition teams overcome the background differences not only in terms of linguistic variations but also in terms of cross-domain knowledge gaps.

We have also presented some example applications using CE and CE-based tools in facilitating cross-linguistic, cross-cultural and cross-domain communication. While we have mostly discussed the general concept and the underlying CE capabilities, we believe that these capabilities can be applied to a wide range of use cases in facilitating cross-domain communication applications[5]. Implementation of applications will be driven by specific use case scenarios.

CE is a simplified and common form of expression in English, which is not only user-friendly in nature but is also restricted in vocabulary and grammar for clear, unambiguous representation and interpretation. The current CE implementation is relatively basic and we are planning to extend it in the areas of syntax, semantics and its general expressivity We believe that modification and/or extension of CE will need to be based not only on theoretical considerations but also on empirical evidence from usability studies and experimentation with real use case scenarios[6].

CE-based communication aid tools are highly interactive. Users can enter information in CE, see reports of information in CE, extend the coverage of CE as a standard language, and get alerted to potential sources of miscommunication in text not authored in CE. To better understand real use case scenarios, more research is needed to allow better modelling of these interactive scenarios in the use of CE-based applications. As we continue to improve the expressivity and naturalness of CE on the one hand and develop more user-friendly functionalities based on assessment of these use cases, on the other, we believe that CE-based communication tools will truly facilitate cross-domain communications and support the cognitive and social processes that enable working together effectively and efficiently.

[5] In fact, CE has been used in various applications. See, for example, Braines et al. (2013) for the discussion of a serviced-based aid for intelligence analysis.
[6] See Xue et al. (2012) for related discussion in a more detailed way.

ACKNOWLEDGMENT

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

Braines, D., Mott, D., Laws, S., (2013) "Controlled English to facilitate human/machine processing", April 2013, SPIE Defense, Security, and Sensing 2013.

Chiarelli, P. W. and P. R. Michaelis (2005) "Winning the peace, the requirement for full-spectrum operations," Military Review, 4-17.

Fellbaum, C. (1998, ed.). WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.

Giammanco, C., McGowan, R., Kao, A., Braines, D., Poteet, S., Pham, T., and Ping, X. (2013) "Knowledge Management for Coalition Information Sharing at the Network Edge," IEEE Intelligent Systems, Vol. 28, No. 1: 26-33.

Miller, G. A. (1995). WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41.

Mott, D. (2009). The representation of logic within semantic web languages, ITACS, url: https://www.usukita.org/papers/5242/details.html

Mott, D. (2010). Summary of Controlled English, ITACS, https://www.usukita.org/papers/5658/details.html.

Mott, D. (2011). The CE linguistic models, ITACS, url: https://www.usukitacs.com/node/2160.

Mott, D., Braines, D., Poteet, S., Kao, A., and Xue, P. (2012). Controlled Natural Language to Facilitate Information Extraction. In Proceedings of the Sixth Annual Conference of the International Technology Alliance, London, UK.

Mott, D., Giammanco, C., Braines, D., Dorneich, M., and Patel, D., (2010). Hybrid Rationale and Controlled Natural Language for Shared Understanding. In Proceedings of the Fourth Annual Conference of the International Technology Alliance, London, UK.

Mott, D and Hendler, J. (2009). Layered Controlled Natural Languages, In Proceedings of the Third Annual Conference of the International Technology Alliance, Maryland, USA.

Mott, D. and Poteet, S. (2012). Ambiguous Words, ITACS, url: https://www.usukitacs.com/node/1907.

Pierce, L. G. (2002a) "Barriers to adaptability in a multinational team," in Proceedings of the 45th Human Factors and Ergonomics Society Annual Meeting, pp. 225-229, Baltimore.

Pierce, L. G. (2002b) "Preparing and supporting adaptable leaders and teams for support and stability operations," in Proceedings of the 11th ROK-US Defense Analysis Seminar (Manpower Policy, Session (4), pp. 97-129, Seoul, Korea.

Poteet, S., Giammanco, C., Patel, J., Kao, A., Xue, P., and Whitely, I. (2009). Miscommunication and context awareness. Proceedings of the 3rd Annual Conference of the International Technology Alliance, Washington.

Poteet, S., Patel, J., Giammanco, C., Whiteley, I, Xue, P., and Kao, A. (2008a). Words are mightier than swords… and yet miscommunication costs lives! Proceedings of the 2rd Annual Conference of the International Technology Alliance, London.

Poteet, S., Xue, P., Patel, J., Kao, A., Giammanco, C. and Whiteley, I. (2008b). Linguistic sources of coalition miscommunication. Paper presented at NATO HFM-142 Symposium, Copenhagen.

Sowa, J. (2007). Common Logic Controlled English, http://www.jfsowa.com/clce/clce07.htm.

Verma, D. (Ed.) (2009). International Technology Alliance in Network and Information Sciences, Biennial Program Plan 2009, Applicable Period: May 12th 2009 - May 11th 2011).

Xue, P., Poteet, S., Kao A., Mott, D., Braines, D., Giammanco, C. and Pham, T. (2012) Information Extraction Using Controlled English to Support Knowledge-Sharing and Decision-Making. The 17th International Command and Control Research and Technology Symposium, Washington DC.

International Technology Alliance
in
Network & Information Sciences

Controlled English for Effective Communication during Coalition Operations

18th ICCRTS
June 19-21, 2013

Stephen Poteet, Ping Xue, Anne Kao
Boeing Research & Technology

David Mott, Dave Braines
IBM UK

Cheryl Giammanco
Army Research Lab

MINISTRY OF DEFENCE

The International Technology Alliance

- Network and Information Sciences International Technology Alliance (ITA) is a collaborative research alliance between the UK Ministry of Defence (UK MoD) and US Army Research Laboratory (US ARL), and a consortium of leading academic and industry partners
- The ITA program started on May 12, 2006; the first phase of the program finished in 2011, and it is now in its second phase (May 2011-May 2016)
- ITA has the strategic goal of producing fundamental advances in information and network sciences
  - to enhance decision making for coalition operations and
  - to enable rapid, secure formation of ad hoc teams in coalition environments and
  - to enhance US and UK capabilities to conduct coalition warfare
- Part of the goal is to address shared understanding and information exploitation in support of decision-making in a coalition environment
- Work presented here is funded under this ITA effort

Challenges for Effective Communication

- Coalition operations involve multi-team and/or multi-nation collaborations
- Various types of linguistic differences exist at various levels of language use between British and American which lead to misunderstanding.
- Cultural differences result in variations in language use even though English speaking nations such as the US and UK share a ‘common’ English language.
- Linguistic variations and cultural differences often create unexpected challenges for effective communication and pose problems for military operations in a coalition setting

Linguistic Variations and Cultural Differences

- **Use of Acronyms**: acronyms usually originate from a specific technical or culture group and are not known by people outside
- **Use of Slang**, **Colloquialisms and Jargon**: slang, colloquial expressions and jargon typically belong to a specific community or group
- **Denotation vs. Connotation**: interpretation of the intended/implied meaning is contextually or culturally dependent
- **Misinterpreted Speech Act**: e.g. British officers give commands such as, "You may well wish to ...", which are interpreted by American officers as suggestions.

Ways to Facilitate Communication

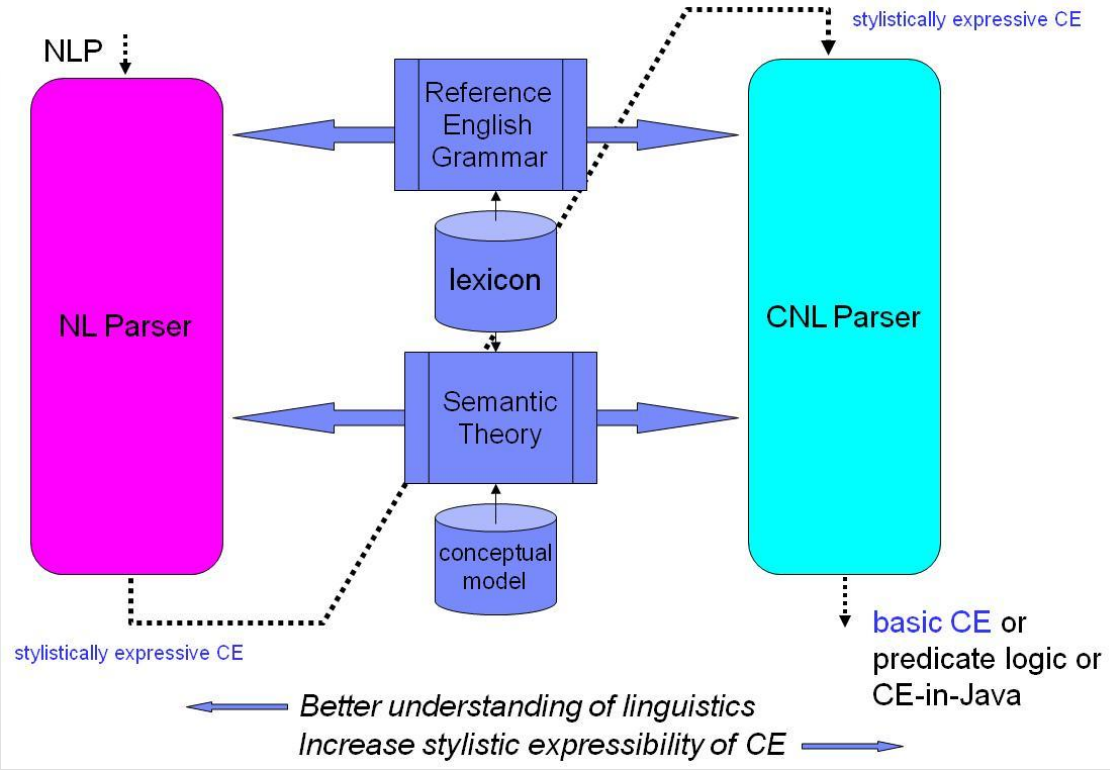
- Have coalition partners train together prior to operations (ideal)
  - However, this might be unrealistic due to time and other constraints
- Promote:
  - Use of a standard terminology and language
  - Supported by automated tools that
    - Enable access to standardized terminology
    - Reinforce consistent language use
    - But also allow extensibility in an ever changing situation
- Controlled English (CE) can help in two ways:
  - A common, extensible standard language, supported and reinforced by CE-based tools, and
  - The basis for communication tools (or add-ons to existing tools) that:
    - Recognize potential sources of miscommunication and
    - Alert users to their presence

ITA Controlled English (CE)

- CE is a type of controlled natural language
- A controlled natural language is a subset of a natural language using a restricted set of grammar rules and a restricted vocabulary
- Traditionally, focus was either on improvement for *human readability* or for *machine readability*
- We are concerned with both easy readability for human consumption and unambiguous representation for computer processing
- Challenge: how to balance user-friendliness and computational predictability

CE is Machine Processable

- ITA CE is consistent with First Order Predicate Logic
  - Based on Common Logic Controlled English (Sowa 2007)
- Syntax is compatible with existing ontology modeling languages such as OWL



Statements and Queries in CE

- CE permits a set of plain English sentences for stating propositions referring to entity existence, properties and relations:
  - ***there is*** a person named Fred.
  - the person Fred ***has*** French as language.
  - the person Fred ***is married to*** the person Jane.

- CE also permits meta statements that specify information about propositions such as their truth status or whether they are assumptions:
  - ***it is true*** *that* there is a person named John.
  - ***it is assumed*** *that* the person Fred is married to the person Jane.

- CE also allows queries:
  - Example: *For which P1 and P2 is it true that the person P1 is the parent of the person P2.*

Inference and Rationale in CE

- CE also allows statements of logical rules to allow inferencing:

  if PREMISES then CONCLUSION

  if (the person X has the person Y as brother) and

  (the person Z has the person X as father)

  then

  (the person Z has the person Y as uncle)

- … and statements that can be used to reconstruct the rationale for an assertion or conclusion:

  CONCLUSION because PREMISES

  the task T1 has the agent A1 as executor

  because

  the plan P1 has the agent A1 as executor and

  the plan P1 contains the task T1.

CE Based Capabilities and Applications

- CE is designed to be easily extensible
- CE encourages a richer interaction and integration between human and machine reasoning capabilities
- CE is most useful in situations that have the following characteristics:
  - A high degree of human interaction, usually involving specialist users with complex needs in non-trivial environments
  - A likelihood of rapidly evolving or uncertain tasks, queries or other knowledge-based activities
  - The need for collaboration, either between different people or teams, and/or across different disciplines

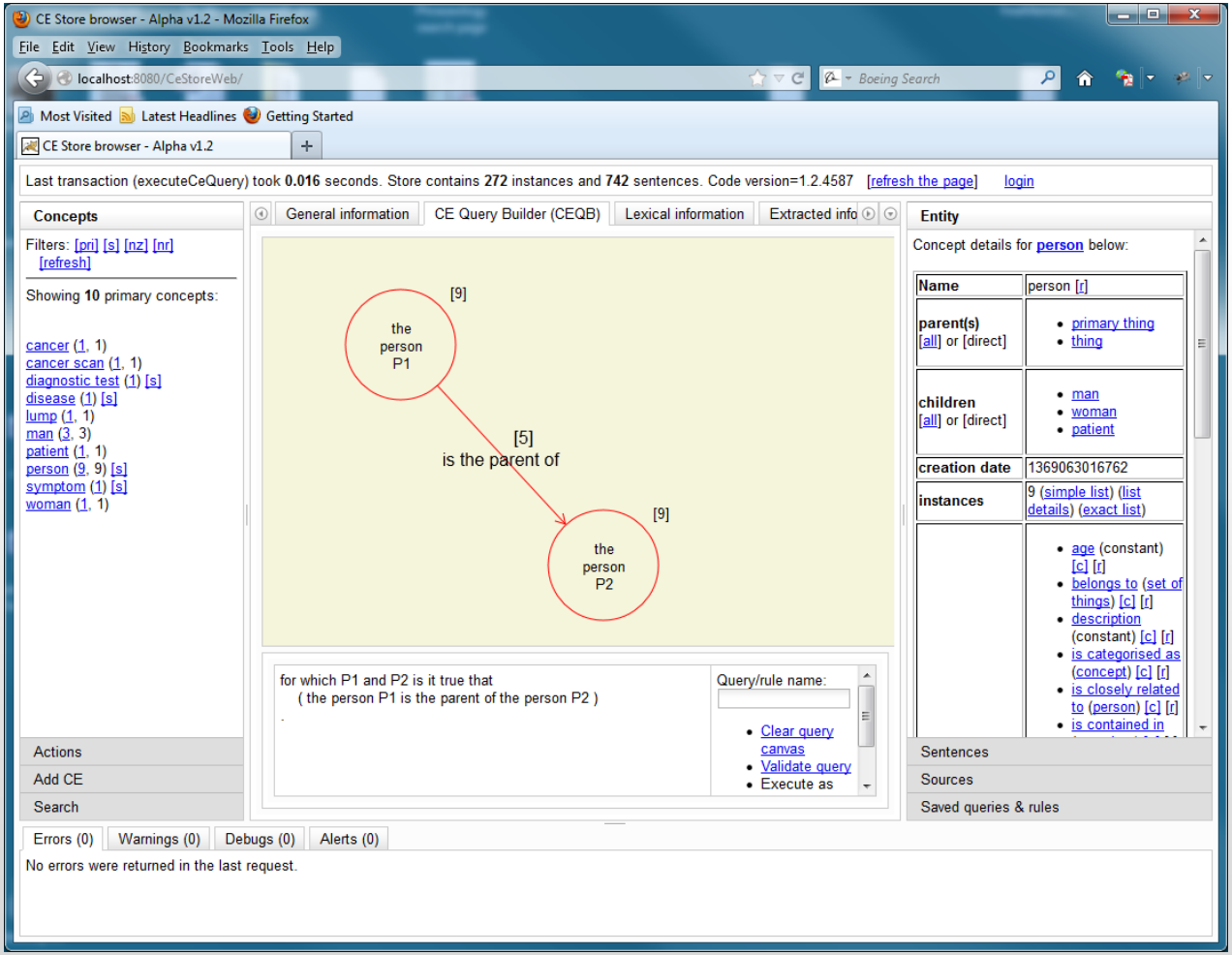
CE Store: A CE-Based Tool Suite

- "CE Store" is a CE-based tool suite to support coalition applications
- It allows one to:
  - Perform basic CE sentence parsing
  - Define and extend any domain concept model
  - Assert any CE sentence conforming to the appropriate conceptual model(s)
  - Define and execute a CE query against a domain model using a combination of a visual query language and written CE
  - Define logical inference rules, in the form of a "query with conclusion clauses"
  - Execute the logical inference rules to infer new CE information and assert it to the knowledge base

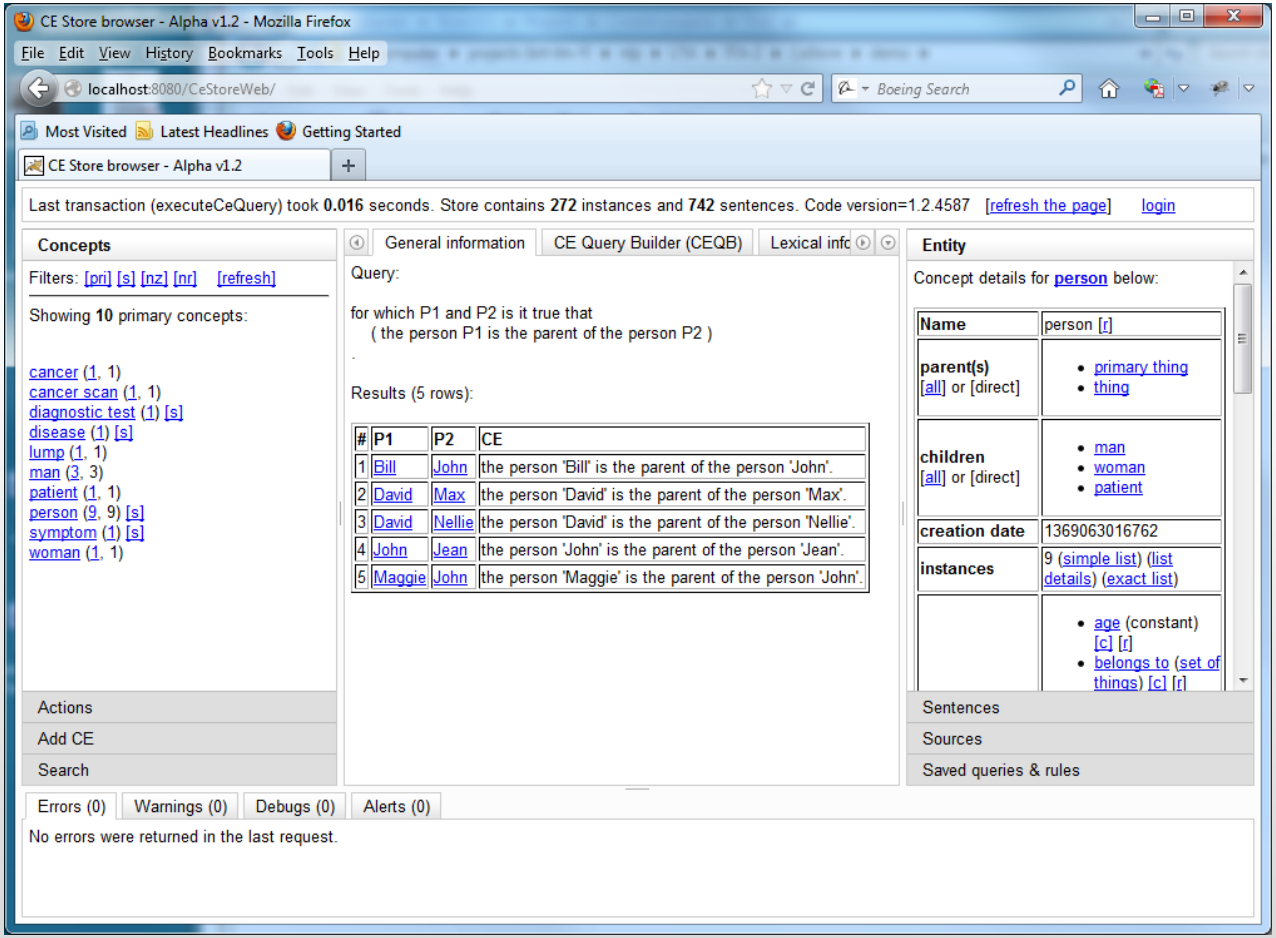
CE Store: CE Query Builder (CEQB)

- A visual query drawing tool embedded in the CE Store environment
  - Uses drag-and-drop and contextual (popup) menus
  - Allows the user to draw, execute and save a CE query or rule
- Visual interface for drawing rules assists user in creating CE rules
- Once constructed, a query (or rule) can then be saved, and executed again in the future
- CEQB is a “model aware” component of the CE Store environment and is directly integrated into the CE Store APIs

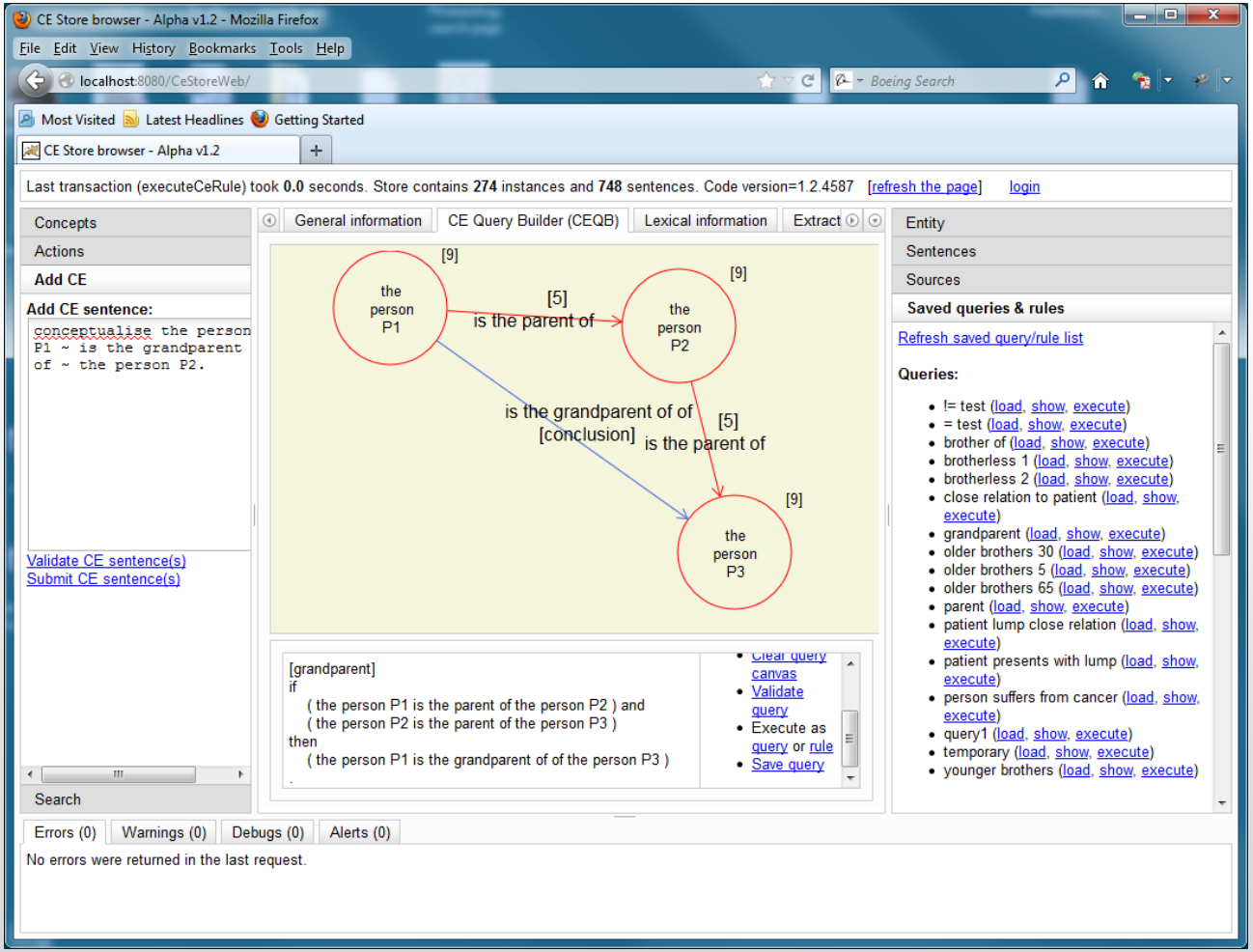
Constructing a Query in CEQB



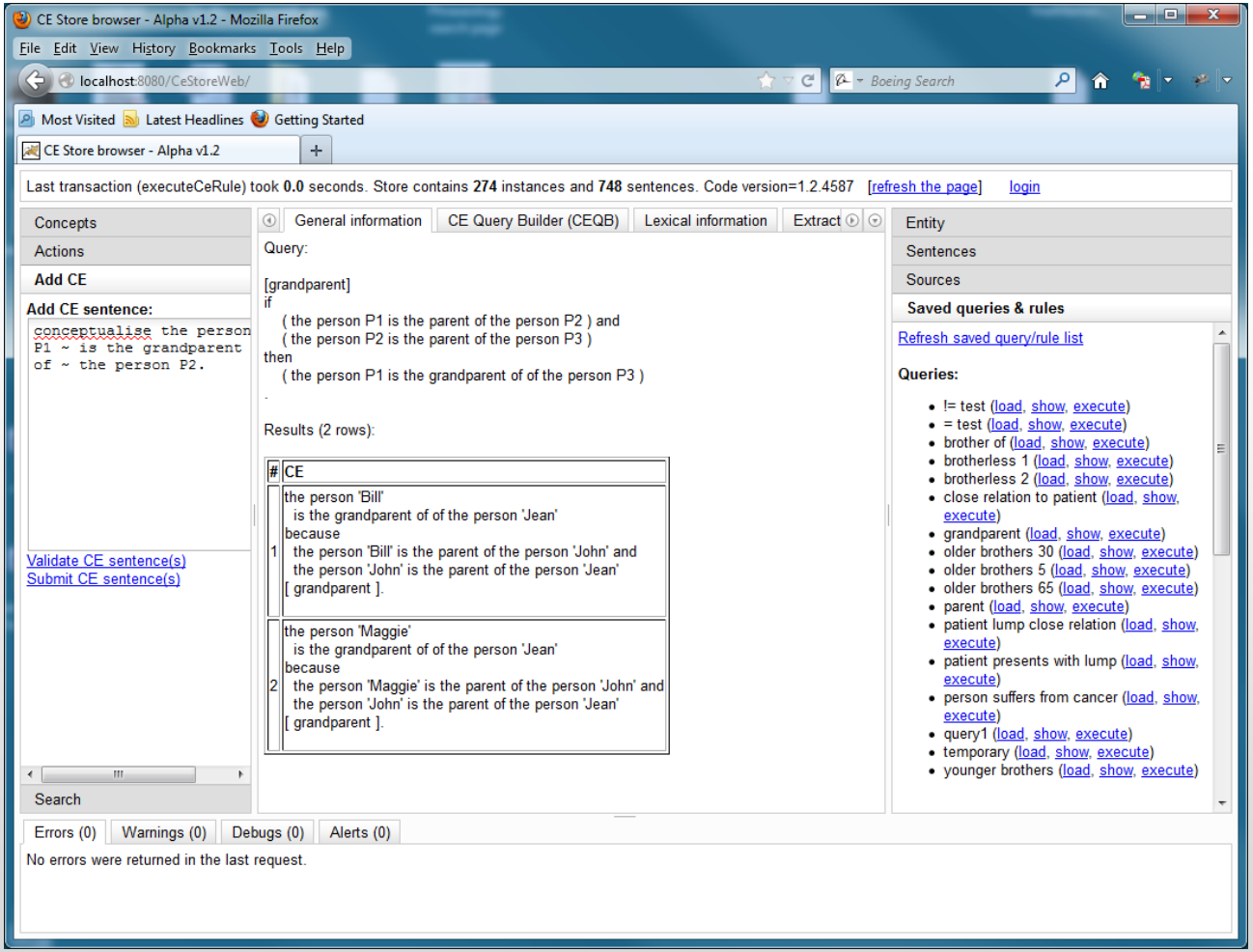
Query Results



Constructing a Rule in CEQB



Rule Execution in CEQB: Rationale



A CE Application

- Facilitating Human/Machine Interaction and Analytical Processing
- Aids the intelligence analysts in a multi-agent collaborative operational environment, especially in cases where the agents are a mixture of other human users and machine processes
- Provides for the relations between objects such as assignment of buildings and vehicles to locations on a map
- Supports complex analytical tasks on synthetic data sources
- Enables teams and team members to have shared situation awareness in a collaboration setting

CE Store Provides an Interactive, Extensible, Multimodal Environment

- **Dynamic model**: Generic concepts are used as starting materials and new concepts are added to the conceptual model by using the following CE sentences:

  conceptualise a ~ building ~ B that is a spatial thing.
  conceptualise a ~ ground feature ~ G that is a spatial thing.

- **Multi-modal interface**: Allows association of photos taken by agents in the field of those objects with icons on the map, and the identification and location of objects extracted from short human generated messages on the map

  there is a building named b1 that has '51.23' as latitude and has '-1.74' as longitude.

- **Live environment:** Enables human users to quickly perform information fusion and adapt their situation awareness in a changing environment

Using CE to Quickly Instantiate a Map

Add CE

Add CE sentence:

```
there is a spatial thing
named 'x' that
  has '51.61721910' as
latitude and
  has '-2.74801318' as
longitude.
```

Validate CE sentence(s)

Submit CE sentence(s)

Benefits of CE and CE-Based Tools

- Provide a simplified and common form of expression in English
- Provide automated tools to enable access to standardized terminology and reinforce consistent language use
- Provide a method to allow extension of standard terminology
- Provide communication aid tools that encourage human-machine interaction, best leveraging human knowledge and computer processing capabilities
- Provide capabilities to construct and extend knowledge models
- Provide an intuitive, CE-based capability for end users to query information from the available information sources

On-going and Future Work

- Apply CE to support Information Extraction
  - As common language to translate into
  - To express the rules of parsing and interpreting
- Extend the CE lexicon by leveraging publicly available lexical-semantic resources such as WordNet
- Extend CE syntax and semantics to extend its general expressivity
- CE extension will be data driven and based on usability studies and experimentation with real use case scenarios
- Develop more user-friendly functionalities based on assessment of real use case scenarios

The International Technology Alliance

ACKNOWLEDGMENT

This research was sponsored by the US Army Research laboratory and the UK Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the US Army Research Laboratory, the U.S. Government, the UK Ministry of Defence, or the UK Government. The US and UK Governments are authorised to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.